



IFW

G&P-105

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of

A. HASHIMOTO

Serial No. 10/807,173

Filed: March 24, 2004

For: COMPUTER SYSTEM, CONTROL APPARATUS,
STORAGE SYSTEM AND COMPUTER DEVICE


TRANSMITTAL OF CERTIFIED PRIORITY DOCUMENT

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Submitted herewith is a certified priority document
(JP 2004-026575, filed February 3, 2004) of a corresponding
Japanese patent application for the purpose of claiming
foreign priority under 35 U.S.C. § 119. An indication that
this document has been safely received would be appreciated.

Respectfully submitted,


Daniel J. Stanger
Registration No. 32,846
Attorney for Applicant

MATTINGLY, STANGER & MALUR
1800 Diagonal Rd., Suite 370
Alexandria, Virginia 22314
(703) 684-1120
Date: June 15, 2004

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 4 年 2 月 3 日
Date of Application:

出 願 番 号 特 願 2 0 0 4 - 0 2 6 5 7 5
Application Number:
[ST. 10/C]: [J . P 2 0 0 4 - 0 2 6 5 7 5]

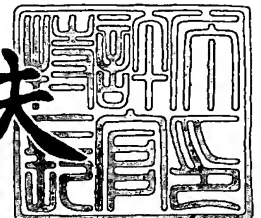
出 願 人 株式会社日立製作所
Applicant(s):

10/807,173
filed 3-24-04
Mattingly, Stanger
& Malus P.C.
703-684-1120
G+P-105

2 0 0 4 年 3 月 5 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



出証番号 出証特 2 0 0 4 - 3 0 1 7 4 6 5

【書類名】 特許願
【整理番号】 GM0311043
【提出日】 平成16年 2月 3日
【あて先】 特許庁長官殿
【国際特許分類】 G06F 12/00
【発明者】
 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所
 システム開発研究所内
 【氏名】 橋本 顕義
【特許出願人】
 【識別番号】 000005108
 【氏名又は名称】 株式会社日立製作所
【代理人】
 【識別番号】 100075513
 【弁理士】
 【氏名又は名称】 後藤 政喜
【選任した代理人】
 【識別番号】 100084537
 【弁理士】
 【氏名又は名称】 松田 嘉夫
【選任した代理人】
 【識別番号】 100114236
 【弁理士】
 【氏名又は名称】 藤井 正弘
【手数料の表示】
 【予納台帳番号】 019839
 【納付金額】 21,000円
【提出物件の目録】
 【物件名】 特許請求の範囲 1
 【物件名】 明細書 1
 【物件名】 図面 1
 【物件名】 要約書 1
 【包括委任状番号】 0110326

【書類名】 特許請求の範囲**【請求項 1】**

アプリケーションソフトウェアが動作するサーバ装置と、前記サーバ装置の動作に必要なデータを記憶するストレージ装置と、を備える計算機システムであって；

前記サーバ装置は、

CPU、メモリ及びI/Oアダプタを含む計算資源と、

前記計算資源を論理的に分割し、独立した仮想計算機として動作させる第1ハイパーバイザと、を有し；

前記ストレージ装置は、

CPU、ディスクキャッシュ、I/Oアダプタ及び物理ディスクを含むストレージ資源と、

前記ストレージ資源を論理的に分割し、独立した仮想ストレージ装置として動作させる第2ハイパーバイザと、を有し；

前記サーバ装置に備わる計算資源を管理するサーバ資源管理テーブルと、

前記ストレージ装置に備わるストレージ資源を管理するストレージ資源管理テーブルと

、
前記仮想計算機と前記仮想ストレージ装置との対応関係を管理する仮想ディスク管理テーブルと、を有する管理部を備え；

前記第1ハイパーバイザは、前記サーバ資源管理テーブルの設定に基づいて前記計算資源を論理的に分割し；

前記第2ハイパーバイザは、前記ストレージ資源管理テーブルの設定に基づいて前記ストレージ資源を論理的に分割することを特徴とする計算機システム。

【請求項 2】

アプリケーションソフトウェアが動作するコンピュータ装置と、前記コンピュータ装置の動作に必要なデータを記憶するストレージ装置とを備える計算機システムであって；

前記コンピュータ装置は、該コンピュータ装置に備わる計算資源を論理的に分割し、独立した仮想計算機として動作させる第1制御部を有し；

前記ストレージ装置は、該ストレージ装置に備わるストレージ資源を論理的に分割し、独立した仮想ストレージ装置として動作させる第2制御部を有し；

前記コンピュータ装置に備わる計算資源を管理する第1管理テーブルと、

前記ストレージ装置に備わるストレージ資源を管理する第2管理テーブルと、

前記仮想計算機と前記仮想ストレージ装置との関係を管理する第3管理テーブルと、を有する管理部を備え；

前記第1制御部は、前記第1管理テーブルの設定に基づいて前記計算資源を論理的に分割し；

前記第2制御部は、前記第2管理テーブルの設定に基づいて前記ストレージ資源を論理的に分割することを特徴とする計算機システム。

【請求項 3】

前記コンピュータ装置は、複数台備わることを特徴とする請求項2に記載の計算機システム。

【請求項 4】

前記ストレージ装置は、複数台備わることを特徴とする請求項2に記載の計算機システム。

【請求項 5】

前記コンピュータ装置には、前記仮想計算機が複数構成され、

前記ストレージ装置には、前記仮想ストレージ装置が複数構成されることを特徴とする請求項2に記載の計算機システム。

【請求項 6】

前記コンピュータ装置が、前記管理部を備え、

前記管理部が、前記第2管理テーブル及び前記第3管理テーブルの設定内容を前記スト

レージ装置に送信することによって、前記ストレージ装置に備わるストレージ資源を論理的に分割し、独立した仮想ストレージ装置として動作させることを特徴とする請求項2に記載の計算機システム。

【請求項7】

前記ストレージ装置が、前記管理部を備え、

前記管理部が、前記第1管理テーブルの設定内容を前記コンピュータ装置に送信することによって、前記コンピュータ装置に備わる計算資源を論理的に分割し、独立した仮想計算機として動作させることを特徴とする請求項2に記載の計算機システム。

【請求項8】

前記コンピュータ装置及び前記ストレージ装置に接続され、前記管理部を備える管理端末を備え、

前記管理部が、

前記第1管理テーブルの設定内容を前記コンピュータ装置に送信することによって、前記コンピュータ装置に備わる計算資源を論理的に分割し、独立した仮想計算機として動作させ、

前記第2管理テーブル及び前記第3管理テーブルの設定内容を前記ストレージ装置に送信することによって、前記ストレージ装置に備わるストレージ資源を論理的に分割し、独立した仮想ストレージ装置として動作させることを特徴とする請求項2に記載の計算機システム。

【請求項9】

前記管理端末は、

前記仮想計算機に割り当てられた計算資源、該仮想計算機に対応する前記仮想ストレージ装置に割り当てられたストレージ資源、及び、計算機システムに備わる計算資源及びストレージ資源を表示し、

前記計算機システムに備わる計算資源及びストレージ資源の中から、各前記仮想計算機に割り当てる計算資源、及び、各前記仮想ストレージ装置に割り当てるストレージ資源の設定を促す表示をすることを特徴とする請求項8に記載の計算機システム。

【請求項10】

前記管理部は、前記コンピュータ装置及び前記ストレージ装置に必要な性能の要求を受け付けると、該性能の実現に必要な前記計算資源及び前記ストレージ資源を算出し、前記第1管理テーブル、前記第2管理テーブル及び前記第3管理テーブルを設定することを特徴とする請求項2に記載の計算機システム。

【請求項11】

アプリケーションソフトウェアが動作するコンピュータ装置と、前記コンピュータ装置の動作に必要なデータを記憶するストレージ装置とに接続される管理端末であって、

前記コンピュータ装置に備わる計算資源を管理する第1管理テーブルと、

前記ストレージ装置に備わるストレージ資源を管理する第2管理テーブルと、

前記仮想計算機と前記仮想ストレージ装置との関係を管理する第3管理テーブルと、を有し、

前記第1管理テーブルの設定内容を前記コンピュータ装置に送信することによって、前記コンピュータ装置に備わる計算資源を論理的に分割し、独立した仮想計算機として動作させ、

前記第2管理テーブル及び前記第3管理テーブルの設定内容を前記ストレージ装置に送信することによって、前記ストレージ装置に備わるストレージ資源を論理的に分割し、独立した仮想ストレージ装置として動作させることを特徴とする管理端末。

【請求項12】

アプリケーションソフトウェアが動作するコンピュータ装置に接続され、前記コンピュータ装置の動作に必要なデータを記憶するストレージ装置であって、

該ストレージ装置に備わるストレージ資源を論理的に分割し、独立した仮想ストレージ装置として動作させる制御部と；

前記コンピュータ装置に備わる計算資源を管理する第1管理テーブルと、
前記ストレージ装置に備わるストレージ資源を管理する第2管理テーブルと、
前記仮想計算機と前記仮想ストレージ装置との関係を管理する第3管理テーブルと、を
有する管理部と、を備え；

前記制御部は、前記第2管理テーブルの設定に基づいて前記ストレージ資源を論理的に
分割し；

前記管理部は、前記第1管理テーブルの設定内容を前記コンピュータ装置に送信するこ
とによって、前記コンピュータ装置に備わる計算資源を論理的に分割し、独立した仮想計
算機として動作させることを特徴とするストレージ装置。

【請求項13】

コンピュータ装置の動作に必要なデータを記憶するストレージ装置に接続され、アプリ
ケーションソフトウェアが動作するコンピュータ装置であって；

該コンピュータ装置に備わる計算資源を論理的に分割し、独立した仮想計算機として動
作させる制御部と；

前記コンピュータ装置に備わる計算資源を管理する第1管理テーブルと、
前記ストレージ装置に備わるストレージ資源を管理する第2管理テーブルと、
前記仮想計算機と前記仮想ストレージ装置との関係を管理する第3管理テーブルと、を
有する管理部と、を備え；

前記制御部は、前記第1管理テーブルの設定に基づいて前記計算資源を論理的に分割し
；

前記管理部は、前記第2管理テーブル及び前記第3管理テーブルの設定内容を前記スト
レージ装置に送信することによって、前記ストレージ装置に備わるストレージ資源を論理
的に分割し、独立した仮想ストレージ装置として動作させることを特徴とするコンピュ
ータ装置。

【書類名】明細書**【発明の名称】** 計算機システム、管理装置、ストレージ装置及びコンピュータ装置**【技術分野】****【0001】**

本発明は、計算機システムに関し、特に、ストレージ装置が接続された計算機システムの記憶装置まで含めた論理分割技術に関する。

【背景技術】**【0002】**

情報処理システムの能力を向上させるための方法として、情報処理システムが有する計算機の台数を増やす方法がある。しかし、計算機を多数設置すると個々の計算機の管理に手間がかかり、またこれらの複数の計算機の設置面積や電力等の合計も大きくなる問題があった。これを解決するために、処理能力の大きい計算機を設置し、その計算機の資源を論理的に複数に分割（L P A R : Logical Partitioning）し、論理分割された仮想的な計算機の各々を独立に使用する技術が提案されている。このような論理分割によって、1台の計算機を仮想的に複数台の計算機に見せることができ、各論理区画にプロセッサ、メモリ等の資源の割り当てを制御することによって、各仮想計算機毎の性能が保証される。また、各仮想的計算機に種々のオペレーティングシステムを自由に搭載し、運用・停止や障害処理も仮想計算機毎に独立して行えるなど、柔軟な運用が可能となる。また、物理的な装置台数が少なくなるので、装置管理、設置面積、消費電力等で有利である。このような論理分割技術は、例えば特許文献1に開示されている。

【特許文献1】 特開 2003-157177号公報

【発明の開示】**【発明が解決しようとする課題】****【0003】**

従来の計算機における論理分割技術では、計算機内のプロセッサ、メモリ等の資源が論理分割されてそれぞれの仮想的な計算機に割り付けられていた。

【0004】

計算機と共に使用される記憶装置には、1台の計算機（ホスト）に直結して用いられるものの他に、ネットワークを介して複数の計算機に共有される記憶装置がある。計算機に接続される記憶装置は、その記憶装置が有する記憶領域が分割され、仮想的な計算機の各々に該分割された記憶領域が割り当てられるだけであった。

【0005】

また、記憶装置にファイルシステム機能を持つことによって、異種サーバ間のファイル共有を可能とした記憶装置、すなわち計算機からファイルアクセスが可能な記憶装置（N A S : Network Attached Storage）が利用されている。N A S とホストとの間のデータのやり取りは、ホスト上で動作するオペレーティングシステムが認識する名前や構造を持ったファイルという単位で行われる。このため、N A S は、データを記憶するためのディスクドライブやその制御部に加え、ホストとのファイル入出力をディスクドライブとのデータ入出力に変換するファイルシステムが動作するプロセッサやメモリを有する。このような N A S は、資源の論理分割が考慮されていなかった。

【0006】

また、大型の外付けストレージ装置に用いられる R A I D（Redundant Array of Independent Disks）システムは、論理分割を前提としていなかった。このような R A I D システムにおいてストレージの論理分割を許容しても、サーバ装置はあらかじめ与えられたストレージ資源の範囲内での論理分割をすることになり、サーバ装置からストレージ装置の資源の再配分をすることは考慮されておらず、サーバとストレージを含めたシステム全体としての資源配分を最適化することはできなかった。

【0007】

本発明は、複数のホストによって共有されるストレージ装置の利用率を高め、ホストとストレージを含めたシステム全体での性能の最適化を図ることを目的とする。

【課題を解決するための手段】

【0008】

アプリケーションソフトウェアが動作するコンピュータ装置と、前記コンピュータ装置の動作に必要なデータを記憶するストレージ装置とを備える計算機システムであって；前記コンピュータ装置は、該コンピュータ装置に備わる計算資源を論理的に分割し、独立した仮想計算機として動作させる第1制御部を有し；前記ストレージ装置は、該ストレージ装置に備わるストレージ資源を論理的に分割し、独立した仮想ストレージ装置として動作させる第2制御部を有し；前記コンピュータ装置に備わる計算資源を管理する第1管理テーブルと、前記ストレージ装置に備わるストレージ資源を管理する第2管理テーブルと、前記仮想計算機と前記仮想ストレージ装置との関係を管理する第3管理テーブルと、を有する管理部を備え；前記第1制御部は、前記第1管理テーブルの設定に基づいて前記計算資源を論理的に分割し；前記第2制御部は、前記第2管理テーブルの設定に基づいて前記ストレージ資源を論理的に分割する。

【発明の効果】

【0009】

本発明によると、サーバの論理分割に対応して、ストレージに論理区画を生成することができるため、サーバとストレージを含めたシステムの資源を最適に配分することができる。

【0010】

また、ディスク以外のストレージ資源（例えば、ディスクキャッシュ）の状態は、サーバから確認することができなかったが、本発明によると、このような性能に顕著に影響する資源も配分をすることができるので、計算機システムの資源配分を最適化することができる。

【発明を実施するための最良の形態】

【0011】

以下、本発明の実施の形態を図面を参照して説明する。

【0012】

図1は、本発明の第1の実施の形態の計算機システムの構成を示すブロック図である。

【0013】

第1の実施の形態の計算機システムは、アプリケーションソフトウェアが動作しているサーバ装置100、サーバ装置100の動作に必要なデータを記憶しているストレージ装置200、及び計算機システム全体の動作を管理する管理端末300によって構成されている。

【0014】

サーバ装置100には、CPU111、メモリ112、I/Oバス113、I/Oアダプタ114、115等の資源を備えた物理計算機110が設けられている。CPU111は、サーバ装置100で実行されるOS(0)132、142及びアプリケーションプログラム133、143に関する演算を行う。メモリ112は、CPU111の動作に必要なプログラムやデータを一時的に保管する。I/Oバス113は、CPU111とI/Oアダプタ114、115との間を接続して、データをやりとりする。I/Oアダプタ114は、I/Oチャネル（例えば、ファイバチャネル）400を介してストレージ装置200と接続されており、ストレージ装置200に対してデータ入出力要求を送信し、ストレージ装置200に記憶されたデータを受け取る。また、I/Oアダプタ115は、ネットワーク410（例えば、イーサネット（登録商標、以下同じ））を介して管理端末300と接続されている。

【0015】

サーバ装置100では、複数のOS132、142が動作している。OS(0)132上では、アプリケーションソフトウェア133が動作しており、OS(1)142上では、アプリケーションソフトウェア143が動作している。このアプリケーションソフトウェア133、143によって、サーバ装置100に接続されたクライアント端末（図示省

略)にデータベースサービスやウェブサービス等が提供される。

【0016】

物理計算機110に備わる資源は、ハイパーバイザ120によって管理されている。ハイパーバイザ120は、サーバ装置100内に論理区画を構成し、独立して動作する仮想計算機を生成する管理ソフトウェアであり、CPU111において実行されることによって、ハイパーバイザ120による処理が実現される。このハイパーバイザ120によって、OS(0)132が使用する計算資源による仮想計算機(0)131と、OS(1)142が使用する計算資源による仮想計算機(1)141とが、物理計算機110上に構成される。

【0017】

ハイパーバイザ120は、仮想ディスク管理テーブル121(図2)を有している。仮想ディスク管理テーブル121には、ストレージ装置200に設けられた仮想ディスク管理テーブル221と同じ内容が記憶されており、ストレージ装置200の仮想ストレージ装置230、240の構成に関する情報が記載されている。

【0018】

ストレージ装置200には、物理ストレージ制御部211、及び物理ディスク215等の資源を備えた物理ストレージ装置210が設けられている。

【0019】

物理ストレージ制御部211は、制御用CPU212、I/Oアダプタ213及びディスクキャッシュ214を有する。制御用CPU212は、物理ディスク215に対するデータの入出力を制御し、ストレージ装置200の動作を管理する。なお、ストレージ装置200がNAS(Network Attached Storage)である場合には、制御用CPU212においてファイルシステムを動作させる。I/Oアダプタ213は、I/Oチャネル400を介してサーバ装置100と接続されている。ディスクキャッシュ214は、物理ディスク215から読み出されるデータや、物理ディスク215に書き込まれるデータを一時的に保管して、サーバ装置100に対するストレージ装置200のアクセス性能を向上させる。

【0020】

物理ディスク215は、ストレージハイパーバイザ220によって管理されている。ストレージハイパーバイザ220は、ストレージ装置200内に論理区画を構成し、独立して動作する仮想ストレージ装置を生成する管理ソフトウェアであり、CPU212において実行されることによって、ストレージハイパーバイザ120による処理が実現される。このストレージハイパーバイザ220によって、仮想ディスク225が構成される。すなわち、ストレージハイパーバイザ220によって、物理ディスク215の領域が分割され複数の仮想ディスク225となったり、複数の物理ディスク215が統合され単一の仮想ディスク225となる。

【0021】

そして、ストレージ装置200は、仮想ディスク225のうち、一つ又は複数の仮想ディスク225を選択し、仮想計算機131、141に記憶領域として提供する。この選択された仮想ディスクを論理ユニットという。論理ユニットとは、OSが一つのディスクとして認識できる単位を指す。

【0022】

また、この論理ユニットにはRAID(Redundant Array of Independent Disks)が構成されており、記憶されるデータに冗長性を持たせている。このため、物理ディスク215の一部に障害が生じて、記憶されたデータが消失しないようになっている。

【0023】

仮想ディスク225に設定された論理ユニットは、仮想ストレージ装置(0)の論理ユニット231と、仮想ストレージ装置(1)の論理ユニット241とに区分されている。そして仮想ストレージ装置(0)が仮想計算機(0)131によってアクセスされ、仮想ストレージ装置(1)が仮想計算機(1)141によってアクセスされる。

【0024】

ストレージハイパーバイザ220は、仮想ディスク管理テーブル221、ディスクアドレス変換テーブル222及びストレージ資源管理テーブル223を有している。

【0025】

仮想ディスク管理テーブル221（図2）には、管理端末300に設けられた仮想ディスク管理テーブル223と同じ内容が記憶されている。

【0026】

ディスクアドレス変換テーブル222（図3）には、仮想ディスクと物理ディスクの対応関係、及び、仮想ディスクのアドレスと物理ディスクのアドレスとの対応関係が規定されている。ディスクアドレス変換テーブル222によって、仮想ディスクのアドレスと物理ディスクのアドレスとが変換される。

【0027】

ストレージ資源管理テーブル223には、管理端末300に設けられたストレージ資源管理テーブル323と同じ内容が記憶されている。

【0028】

管理端末300は、計算機システムを統括して管理するコンピュータ装置で、仮想計算機管理プログラム310が動作している。仮想計算機管理プログラム310は、仮想ディスク管理テーブル321、ストレージ資源管理テーブル323及びサーバ資源管理テーブル324を有している。

【0029】

仮想ディスク管理テーブル321には、ストレージ装置200に設けられた仮想ディスク管理テーブル221と同じ内容が記憶されている。

【0030】

ストレージ資源管理テーブル323（図4）には、ストレージ装置200の資源と仮想計算機との対応関係が規定されている。ストレージ資源管理テーブル223によって、ストレージ資源の配分が管理される。

【0031】

サーバ資源管理テーブル324（図5）は、サーバ装置100の資源と仮想計算機との対応関係が規定されている。サーバ資源管理テーブル324によって、サーバ装置100の計算機資源が管理される。

【0032】

また、管理端末300は、サーバ装置100及びストレージ装置200と、ネットワーク410によって接続されている。サーバ装置100、ストレージ装置200及び管理端末300は、ネットワーク410を介して計算機システムの管理情報（各管理テーブルの内容）を送受信している。

【0033】

具体的には、仮想ディスク管理テーブル321は、仮想計算機管理プログラム310によって生成され、ストレージ装置200に転送されて仮想ディスク管理テーブル221となる。この仮想ディスク管理テーブル321には、仮想計算機に対応する仮想ストレージ装置の構成が規定されている。仮想ディスク管理テーブル321によって、どの仮想計算機がどの論理ユニットにアクセス可能なかが管理される。

【0034】

また、ストレージ資源管理テーブル323は、仮想計算機管理プログラム310によって生成され、ストレージ装置200に転送されてストレージ資源管理テーブル223となる。また、これらのテーブルの更新情報もネットワーク410を介して送受信される。

【0035】

I/Oチャネル400は、例えばファイバチャネルプロトコルのような、データの転送に適するプロトコルで通信可能な伝送媒体である。なお、サーバ装置100とストレージ装置200との間は、1対1で接続されても、ネットワーク（SAN）によって接続されてもよい。

【0036】

ネットワーク 410 は、例えば、TCP/IP プロトコルによって、コンピュータ間でデータや制御情報を通信可能に構成されており、例えばイーサネットが用いられる。

【0037】

以上説明した第 1 の実施の形態では、1 台のサーバ装置 100 と 1 台のストレージ装置 200 とが接続されている例について説明したが、サーバ装置 100、ストレージ装置 200 の一方又は双方が複数台設けられてもよい。

【0038】

また、仮想計算機と仮想ストレージとが一対一対応で接続されている例について説明したが、複数の仮想計算機と一つの仮想ストレージとが対応するように接続されても、一つの仮想計算機と複数の仮想ストレージとが対応するように接続されてもよい。

【0039】

図 2 は、本発明の実施の形態の仮想ディスク管理テーブルの説明図である。

【0040】

前述したように、仮想ディスク管理テーブル 221 は、ユーザが管理端末 300 を操作することによって、管理端末 300 内に生成され、同じ内容のテーブルが、サーバ装置 100 に仮想ディスク管理テーブル 121 として記憶され、ストレージ装置 200 に仮想ディスク管理テーブル 221 として記憶される。

【0041】

仮想ディスク管理テーブル 221 には、仮想計算機番号 401、論理ユニット番号 402 及び仮想ディスク番号 403 が対応するように格納されている。仮想計算機番号 401 は、サーバ装置 100 に設けられた仮想計算機に対応する。論理ユニット番号 402 は、仮想ディスク番号 403 に記載された仮想ディスク 225 によって構成された論理ユニットに付されている番号である。

【0042】

この仮想ディスク管理テーブル 221 によって、どの仮想計算機がどの論理ユニットにアクセス可能なのか（すなわち、どの仮想ディスクにアクセス可能なのか）を知ることができる。

【0043】

図 3 は、本発明の実施の形態のディスクアドレス変換テーブルの説明図である。

【0044】

前述したように、ディスクアドレス変換テーブル 222 は、ストレージハイパーバイザ 220 によってストレージ装置 200 内に生成され、仮想ディスクと物理ディスクの対応関係、及び、仮想ディスクのアドレスと物理ディスクのアドレスとの対応関係を記憶している。

【0045】

ディスクアドレス変換テーブル 222 には、仮想ディスク番号 501、仮想ブロックアドレス 502、物理ディスク番号 503 及び物理ブロックアドレス 504 が対応するように格納されている。仮想ディスク番号 501 は、ストレージハイパーバイザ 220 によって生成された仮想ディスク 225 の番号であり、仮想ディスク管理テーブルに格納された仮想ディスク番号 403 と対応している。仮想ブロックアドレス 502 は、仮想ディスク 225 のアドレスが記載されている。仮想ブロックアドレス 502 は、物理ディスク番号 503 に規定された物理ディスク 215 の物理ブロックアドレス 504 に対応している。すなわち、121 番の仮想ディスクのアドレス 0x00000000 は、8 番の物理ディスクの 0x00000000 に対応している。また、121 番の仮想ディスクのアドレス 0x80000000 は 9 番の物理ディスクの 0x00000000 に対応している。つまり、121 番の仮想ディスクは、8 番の物理ディスクと 9 番の物理ディスクとによって構成されている。また、ディスクアドレス変換テーブル 222 によって仮想ディスクのアドレスと物理ディスクのアドレスとを変換することができる。

【0046】

図4は、本発明の実施の形態のストレージ資源管理テーブルの説明図である。

【0047】

前述したように、ストレージ資源管理テーブル323は、ユーザが管理端末300を操作することによって、管理端末300内に生成され、同じ内容のテーブルが、ストレージ装置200にストレージ資源管理テーブル223として記憶される。

【0048】

なお、後述する第2の実施の形態(図13)では、ストレージ資源管理テーブル223としてストレージ装置200内に生成される。また、後述する第3の実施の形態(図14)では、ストレージ資源管理テーブル223としてサーバ装置100内に生成される。

【0049】

ストレージ資源管理テーブル323には、仮想計算機番号601、仮想ディスク番号602、ディスクキャッシュ容量603、制御CPU番号604及びI/Oアダプタ番号605が対応するように格納されている。ストレージ資源管理テーブル323は、ストレージ装置200の資源(仮想ディスク225、制御CPU212、I/Oアダプタ213、ディスクキャッシュ214)と仮想計算機との対応関係を記憶している。

【0050】

仮想計算機番号601は、サーバ装置100に設けられた仮想計算機に対応する。仮想ディスク番号602は、ストレージハイパーバイザ220によって構成された仮想ディスク225の番号であり、仮想計算機番号601に規定された仮想計算機に割り当てられた仮想ディスクを示す。この仮想ディスク番号602は、仮想ディスク管理テーブル221に格納された仮想ディスク番号403と対応している。

【0051】

ディスクキャッシュ容量603は、仮想計算機番号601に規定された仮想計算機に割り当てられたディスクキャッシュ214の容量である。制御CPU番号604は、仮想計算機番号601に規定された仮想計算機からのアクセス(仮想ディスク番号602に規定された仮想ディスクへのアクセス)を制御する制御用CPU212を示す。

【0052】

I/Oアダプタ番号605は、仮想計算機番号601に規定された仮想計算機からのアクセス(仮想ディスク番号602に規定された仮想ディスクへのアクセス)を受け持つI/Oアダプタ213を示す。

【0053】

すなわち、仮想計算機(0)131には、121番～123番の3個の仮想ディスク225が割り当てられている。この121番～123番の仮想ディスク225に対してアクセスするために仮想計算機(0)131は512Mバイトのディスクキャッシュを使用することができる。仮想計算機(0)131から121番～123番の仮想ディスク225に対するアクセスは0番～2番の3個のI/Oアダプタを介して行われる。仮想計算機(0)131から121番～123番の仮想ディスク225に対するアクセスを処理するために48番～50番の3個のCPUが動作することを意味している。

【0054】

図5は、本発明の実施の形態のサーバ資源管理テーブルの説明図である。

【0055】

前述したように、資源管理テーブルは、第1の実施の形態では仮想計算機管理プログラム310によって管理端末300内にサーバ資源管理テーブル324として作られる。

【0056】

なお、後述する第2の実施の形態(図13)では、サーバ資源管理テーブル224としてストレージ装置200内に生成される。また、後述する第3の実施の形態(図14)では、サーバ資源管理テーブル123としてサーバ装置100内に生成される。

【0057】

資源管理テーブルには、仮想計算機番号701、CPU配分率702、メモリ容量703及びI/Oアダプタ番号704が対応するように格納されている。資源管理テーブル3

24は、サーバ装置100の資源（CPU111、メモリ112、I/Oアダプタ114）の対応関係を記憶している。

【0058】

仮想計算機番号701は、サーバ装置100に設けられた仮想計算機を示す。CPU配分率702は、サーバ装置100に設けられたCPUのうち当該仮想計算機に割り当てられた割合を示す。メモリ容量703は、当該仮想計算機に割り当てられたメモリ112の容量である。I/Oアダプタ番号704は、当該仮想計算機からストレージ装置200へのアクセスを受け持つI/Oアダプタ114を示す。

【0059】

図6は、本発明の実施の形態の資源割り当て設定処理のフローチャートである。

【0060】

まず、ユーザは管理端末300を操作して、サーバ装置100が有する計算機資源（CPU111、メモリ112、I/Oアダプタ114等）及びストレージ装置200が有する資源（CPU212、I/Oアダプタ213、ディスクキャッシュ214、仮想ディスク225）の、仮想計算機毎の割り当てを設定して、サーバ資源管理テーブル324を更新する（S101）。管理端末300は、設定された資源割り当てをサーバ装置100に送信する（S102）。

【0061】

サーバ装置100は、管理端末300から資源割り当てを受信すると、この設定に従ってサーバ装置100内の計算機資源を割り当て、仮想計算機を生成する（S103）。そして、仮想計算機の生成が終了すると、管理端末300に仮想計算機の設定が完了した旨を報告する（S104）。

【0062】

管理端末300は、サーバ装置100から仮想計算機の設定が完了した旨の報告を受信すると、次にストレージ装置200に設定された資源割り当て（ストレージ資源管理テーブルの更新情報）を送信する（S105）。

【0063】

ストレージ装置200は、管理端末300から資源割り当てを受信すると、この設定に従ってストレージ資源管理テーブル223及び仮想ディスク管理テーブル221を更新して、ストレージ装置200内の資源を割り当てる（S106）。また、必要であれば、仮想ディスク管理テーブル221及びディスクアドレス変換テーブル222を更新して、仮想ストレージ装置を生成、又は、その構成を変更する（S106）。そして、仮想ストレージ装置の生成が終了すると、管理端末300に仮想ストレージ装置の設定が完了した旨を報告する（S107）。

【0064】

図7は、ストレージ装置200におけるデータ入出力処理のフローチャートである。

【0065】

ストレージ装置200は、サーバ装置100から入出力コマンドを受信する（S111）。この入出力コマンドは、ストレージハイパーバイザ220に伝えられる。ストレージハイパーバイザ220は、この入出力コマンド（ハイパーバイザ通信ヘッダ1203。図9、図10参照）に含まれる送信元仮想計算機番号1302と送信先仮想計算機番号1303を読み取る（S112）。そして、ストレージハイパーバイザは送信先仮想計算機番号1303に対応する仮想ストレージ装置にハイパーバイザ通信ペイロード1204を転送する（S113）。本実施の形態では、ハイパーバイザ通信ペイロード1204には、ディスクI/Oコマンドがふくまれており、仮想ストレージ装置はこれを実行することになる。

【0066】

そして、当該アクセスの対象となる仮想ディスク番号を取得して、アクセス先の仮想ディスクを特定して、対応する仮想ディスク225にアクセスする（S114）。

【0067】

仮想ディスク 225 に対するアクセスは、ストレージハイパーバイザ 220 によって受け付けられる。ストレージハイパーバイザ 220 は、ディスクアドレス変換テーブル 222 を用いて、アクセス対象の仮想ディスクの仮想ブロックアドレスに対応する物理ディスクの物理ブロックアドレスを特定し、仮想ディスク 225 に対するアクセスを物理ディスク 215 に対するアクセスに変換する。そして、ストレージハイパーバイザ 220 が物理ディスク 215 にアクセスし、データの読み出し処理又は書き込み処理を行う (S115)。

【0068】

ストレージハイパーバイザ 220 は、物理ディスク 215 に対するデータ入出力処理が終了すると、仮想ストレージ装置に対してデータ入出力処理の結果を報告する (S116)。仮想ストレージ装置は、ストレージハイパーバイザ 220 からのデータ入出力処理の結果を受信すると、ストレージハイパーバイザ 220、ハイパーバイザ 110 を介して、仮想計算機に対してデータ入出力処理の結果を報告する (S117、S118、S119)。

【0069】

次にサーバ装置 100 とストレージ装置 200 の入出力命令の処理を説明する。サーバ装置 100 とストレージ装置 200 の通信は、I/O チャンネル 400 を介して行われる。I/O チャンネル 400 での通信を、一般のファイバチャネル、イーサネットと同様の層構造をなすプロトコルを例にして説明する。

【0070】

図 8 は、I/O チャンネル 400 の通信プロトコル層の構造の説明図である。

【0071】

仮想計算機 (0) 131 上の OS (0) 132 は、ストレージ装置 200 内の論理ユニットにアクセスする場合、ディスク I/O プロトコル (例えば、SCSI 規格) に従って入出力を行う。本実施の形態では、ディスク I/O プロトコル層を「ディスク I/O 層」1100、1106 と称す。OS (0) 132 が発行したディスク I/O コマンドは、ハイパーバイザ 120 によって受け取られるが、ハイパーバイザ 120 とストレージハイパーバイザ 220 との間にも通信プロトコル層が存在する。これを「ハイパーバイザ通信層」1101、1105 と称す。さらに、本実施の形態では、I/O チャンネル 400 による汎用的な通信を行う層を「I/O チャンネルプロトコル層」1102、1104 と称す。さらに、物理媒体などのハードウェアの層を「物理層」1103 と称す。このような層構造をなすことにより、I/O チャンネル 400 の物理媒体が変更しても、ディスク I/O 層 1100、1106 やハイパーバイザ通信層 1101、1105 が影響を受けない。

【0072】

OS (0) 132 が発行したディスク I/O コマンドは、仮想計算機 (0) 131 に伝えられる。仮想計算機 (0) 131 は仮想ストレージ装置 (0) に対して当該 I/O コマンドを発行する。この当該 I/O コマンドは、実際にはハイパーバイザ 120 が受け取る。ハイパーバイザ 120 は、当該ディスク I/O コマンドに情報を付加して (図 9 参照)、ストレージハイパーバイザ 220 に送信する。ストレージハイパーバイザ 220 はこれを受信し、当該ディスク I/O コマンドを抽出して仮想ストレージ装置 (0) に送信する。このような層構造で通信をすることによって、OS (0) 132 は直接仮想ストレージ装置 (0) と通信しているように認識する。

【0073】

図 9 は、サーバ装置 100 とストレージ装置 200 との間で伝達される情報の説明図である。

【0074】

本実施の形態における I/O チャンネル 400 においては、フレーム 1200 を単位として通信が行われる。これは一般のファイバチャネル、イーサネットと同様である。フレーム 1200 は、I/O チャンネルプロトコルヘッダ 1201 及び I/O チャンネルプロトコルペイロード 1202 からなる。I/O チャンネルプロトコルヘッダ 1201 は、I/O チャ

ネルプロトコル層 1102、1104 での通信に必要な制御情報を含んでいる。図示を省略するが、例えば、送信元の識別子や、送信先の識別子である。I/Oチャネルプロトコルペイロード 1202 は、I/Oチャネルプロトコル層 1102、1104 において伝達される情報であり、I/Oチャネルプロトコル層 1102、1104 ではその内容に関知しない。

【0075】

I/Oチャネルプロトコルペイロード 1202 は、ハイパーバイザ通信ヘッダ 1203 及びハイパーバイザ通信ペイロード 1204 からなる。ハイパーバイザ通信ヘッダ 1203 には、ハイパーバイザ通信層 1101、1105 での通信に必要な制御情報を含んでいる（後述）。ハイパーバイザ通信ペイロード 1204 は、ハイパーバイザ通信層 1101、1105 において伝達されるデータであり、ハイパーバイザ通信層 1101、1105 ではその内容には関知しない。

【0076】

本実施の形態では、ハイパーバイザ通信ペイロード 1204 は、ディスク I/O 層 1100、1106 の通信に必要な情報からなる。具体的には、ディスク I/O コマンドや、転送されるデータである。本実施の形態では、ハイパーバイザ通信層 1101、1105 の上位層がディスク I/O 層であるため、ハイパーバイザ通信ペイロード 1204 にディスク I/O 層 1100、1106 の情報が含まれる。しかし、ハイパーバイザ、ストレージハイパーバイザ間の通信であれば、ディスク I/O ではない別の形式の情報が含まれる。

【0077】

図 10 は、ハイパーバイザ通信ヘッダ 1203 の内容の説明図である。

【0078】

ハイパーバイザ通信ヘッダ 1203 は、本発明の実施の形態特有のものであり、送信元ハイパーバイザ番号 1300、送信先ハイパーバイザ番号 1301、送信元仮想計算機番号 1303、送信先仮想計算機番号 1304 からなる。本実施の形態では、サーバ装置 100、ストレージ装置 200 が、それぞれ複数存在するシステムにも対応できるように、ハイパーバイザ、ストレージハイパーバイザに一意な識別子を付与する。

【0079】

送信元ハイパーバイザ番号 1300 は、本フレームを送信したハイパーバイザ又はストレージハイパーバイザの識別子である。

【0080】

送信先ハイパーバイザ番号 1301 は、本フレームを受信すべきハイパーバイザ又はストレージハイパーバイザの識別子である。

【0081】

送信元仮想計算機番号 1302 は、本フレームを送信した仮想計算機、又は仮想ストレージ装置の識別子である。

【0082】

送信先仮想計算機番号 1303 は、本フレームを受信すべき仮想計算機、又は仮想ストレージ装置の識別子である。

【0083】

図 11、図 12 は、本発明の実施の形態の計算機システムの構成の設定画面の説明図である。

【0084】

画面上部には仮想計算機毎に割り当てられた資源を設定するページが設けられている。また、画面下部にはサーバ装置 100 及びストレージ装置 200 に備わる全資源（リソース）の表示欄が設けられている。なお、全資源の他に、利用されていない（又は、既に利用されている）リソースを表示してもよい。

【0085】

管理者は、この画面上部の資源毎のページにサーバ装置やストレージ装置の資源を書き

込むことによって、又は、画面下部の全リソースの表示欄から資源を移動させることによって、仮想計算機毎の資源を設定することができる。

【0086】

また、管理者が個別に仮想計算機と仮想ストレージ装置に割り当てることなく、仮想計算機（仮想ストレージ装置を含む）に要求される性能を設定することによって、該性能の実現に必要な資源を算出し、仮想計算機と仮想ストレージ装置の資源を設定することもできる。

【0087】

例えば、データの読み出し性能を重視する仮想計算機であれば、仮想ストレージ装置に割り当てられるディスクキャッシュ 214 の容量を大きく設定する。また、ディスクキャッシュ 214 の全資源が少なく、仮想ストレージ装置に割り当てられるディスクキャッシュ 214 の容量が小さければ、仮想計算機のメモリの割当を大きく設定する。一方、ディスクキャッシュ 214 の全資源が多く、仮想ストレージ装置に割り当てられるディスクキャッシュ 214 の容量が少なければ、仮想計算機のメモリの割当を小さくする。

【0088】

さらに、仮想計算機で動作しているアプリケーションソフトウェアが、ディスク上の広範囲にランダムアクセスするものであれば、キャッシュの効果を生じることが少ないので、ディスクキャッシュ 214 の割当を小さくする。また、動画データをストリーミング配信する等のマルチメディア系のアプリケーションソフトウェアの場合は、仮想ストレージ装置に割り当てられるディスクキャッシュ 214 の容量を大きくし、仮想計算機に割り当てられるメモリ 112 の容量も大きくする。

【0089】

また、サーバ装置 100 やストレージ装置を増設したり、減少させたりするときに、仮想計算機や仮想ストレージ装置の構成もこの画面によって設定することができる。

【0090】

このように、本発明の第 1 の実施の形態では、サーバ資源管理テーブル 324 と、ストレージ資源管理テーブル 323 と、仮想ディスク管理テーブル 321 とを備え、ハイパーバイザ 120 はサーバ資源管理テーブル 324 の設定に基づいて計算資源を論理的に分割し、独立した仮想計算機として動作させ、ストレージハイパーバイザ 220 は、前記ストレージ資源管理テーブル 323 の設定に基づいて前記ストレージ資源を論理的に分割し、独立した仮想ストレージ装置として動作させるので、サーバ装置とストレージ装置を含めた計算機システムの資源を統合的に管理することができ、これらの資源を最適に配分することができる。

【0091】

また、仮想計算機の構成を変化させるときに、仮想ストレージ装置の構成を変化させることができるので、仮想計算機と仮想ストレージ装置とを別個に設定する必要がなく、計算機システム全体の性能を考慮しながら見ながら仮想計算機と仮想ストレージ装置のリソースを設定することができる。また、従来、管理端末 300 から管理することができなかったディスクキャッシュ 214 のような資源も、仮想計算機の資源と同時に設定することができる。

【0092】

また、本実施の形態では、利用者が、図 11 に示す設定画面から、「ディスク」部分に関する詳細設定を別画面に表示させて設定することができる。言うまでもないが、本発明は画面表示の方法に依存しない。

【0093】

図 12 は、この詳細設定画面の説明図である。

【0094】

詳細設定画面（図 12）は、図 11 に示す設定画面において「詳細」ボタンを押すことによって、仮想計算機毎に設けられた詳細設定画面を呼び出すことができる。本実施例では、論理ユニット 0 は、二つの物理ディスク（物理ディスク 8 番、9 番）から構成されて

いる。この詳細設定画面に物理ディスク毎に表示されている「10,000 r p m」は、物理ディスク 8 番、9 番が、磁気記録媒体である磁気ディスクを毎分 10,000 回転できる仕様であることを示している。磁気ディスクの回転数も、物理ディスクの性能を規定する重要な要素である。使用者は、性能を要求される用途には、高性能の物理ディスクを本画面にて選択し、論理ユニットを構成することができる。また使用者は、一つの論理ユニットを構成する物理ディスクの数を多くして、論理ユニットの性能を上げることができる。

【0095】

以上説明したように、本発明によってストレージの資源を仮想計算機を関連付けて配分することが可能となり、サーバとストレージをあわせた計算機システム全体として最適な資源配分が可能になる。

【0096】

図 13 は、本発明の第 2 の実施の形態の計算機システムの構成を示すブロック図である。

【0097】

第 2 の実施の形態では、前述した第 1 の実施の形態（図 1）と異なり、管理端末 300 を有さず、第 1 の実施の形態における管理端末 300 の機能がストレージ装置 200 に備わっている。なお、第 1 の実施の形態と同じ構成は同じ符号を付し、その詳細な説明を省略する。

【0098】

第 1 の実施の形態の計算機システムは、アプリケーションソフトウェアが動作しているサーバ装置 100、計算機システム全体の動作を管理し、サーバ装置 100 の動作に必要なデータを記憶しているストレージ装置 200、及び、計算機システム全体の動作の指示をストレージ装置 200 にする管理端末 350 によって構成されている。

【0099】

サーバ装置 100 には、CPU 111、メモリ 112、I/O バス 113、I/O アダプタ 114、115 等の資源を備えた物理計算機 110 が設けられている。サーバ装置 100 の構成及び動作は前述した第 1 の実施の形態と同じである。

【0100】

ストレージ装置 200 には、物理ストレージ制御部 211、及び物理ディスク 215 等の資源を備えた物理ストレージ装置 210 が設けられている。

【0101】

ストレージハイパーバイザ 220 は、仮想ディスク管理テーブル 221、ディスクアドレス変換テーブル 222、ストレージ資源管理テーブル 223 及びサーバ資源管理テーブル 224 を有している。

【0102】

仮想ディスク管理テーブル 221（図 2）、ディスクアドレス変換テーブル 222（図 3）、ストレージ資源管理テーブル 223（図 4）は、前述した第 1 の実施の形態と同じである。サーバ資源管理テーブル 224（図 5）は、サーバ装置 100 の資源と仮想計算機との対応関係が規定されている。サーバ資源管理テーブル 224 によって、サーバ装置 100 の計算機資源が管理される。

【0103】

また、ストレージハイパーバイザ 220 は、これらの管理テーブル 221、223、224 を用いて、計算機システムを統括して管理する。

【0104】

また、ストレージハイパーバイザ 220 では、計算機システムを統括して管理する仮想計算機管理プログラムが動作しており、これらの管理テーブル 221、223、224 を用いて、計算機システムを統括して管理する。

【0105】

管理端末 350 は、計算機システムの管理情報を設定するコンピュータ装置で、ストレージ装置 200 と接続されている。すなわち、管理者は管理端末 350 を操作することに

よって、ストレージ資源管理テーブル 223 と、サーバ資源管理テーブル 224 とに設定される内容を更新することができる。

【0106】

このように、本発明の第 2 の実施の形態では、前述した第 1 の実施の形態の効果に加え、管理端末 300 の機能をストレージ装置 200 に備えたので、別に管理端末を設けることなく、仮想計算機に対応した仮想ストレージの管理をすることができる。

【0107】

図 14 は、本発明の第 3 の実施の形態の計算機システムの構成を示すブロック図である。

【0108】

第 3 の実施の形態では、前述した第 1 又は、第 2 の実施の形態（図 1）と異なり、管理端末 300 を有さず、第 1 の実施の形態における管理端末 300 の機能がサーバ装置 100 に備わっている。なお、第 1 の実施の形態と同じ構成は同じ符号を付し、その詳細な説明を省略する。

【0109】

第 3 の実施の形態の計算機システムは、アプリケーションソフトウェアが動作しており、計算機システム全体の動作を管理するサーバ装置 100、及びサーバ装置 100 の動作に必要なデータを記憶しているストレージ装置 200 によって構成されている。

【0110】

サーバ装置 100 には、CPU 111、メモリ 112、I/Oバス 113、I/Oアダプタ 114、115 等の資源を備えた物理計算機 110 が設けられている。これら物理計算機 110 の構成は前述した第 1 の実施の形態と同じである。

【0111】

物理計算機 110 に備わる資源は、ハイパーバイザ 120 によって管理されている。このハイパーバイザ 120 によって、OS (0) 132 が使用する計算資源による仮想計算機 (0) 131 と、OS (1) 142 が使用する計算資源による仮想計算機 (1) 141 とが、物理計算機 110 上に構成される。ハイパーバイザ 120 は、仮想ディスク管理テーブル 121、ストレージ資源管理テーブル 123 及びサーバ資源管理テーブル 124 を有している。

【0112】

仮想ディスク管理テーブル 121 には、ストレージ装置 200 に設けられた仮想ディスク管理テーブル 221 と同じ内容が記憶されている。

【0113】

ストレージ資源管理テーブル 123（図 4）には、ストレージ装置 200 の資源と仮想計算機との対応関係を規定されている。ストレージ資源管理テーブル 223 によって、ストレージ資源の配分が管理される。

【0114】

サーバ資源管理テーブル 124（図 5）は、サーバ装置 100 の資源と仮想計算機との対応関係が規定されている。サーバ資源管理テーブル 224 によって、サーバ装置 100 の計算機資源が管理される。

【0115】

ハイパーバイザ 120 では、計算機システムを統括して管理する仮想計算機管理プログラムが動作しており、これらの管理テーブル 121、123、124 を用いて、計算機システムを統括して管理する。すなわち、管理者はサーバ装置 100 を操作することによって、ストレージ資源管理テーブル 123 と、サーバ資源管理テーブル 124 とに設定される内容を更新することができる。

【0116】

ストレージ装置 200 には、物理ストレージ制御部 211、及び物理ディスク 215 等の資源を備えた物理ストレージ装置 210 が設けられている。これらストレージ装置 200 の構成は前述した第 1 の実施の形態と同じである。なお、ストレージ資源管理テーブル

223には、サーバ装置100に設けられたストレージ資源管理テーブル123と同じ内容が記憶されている。

【0117】

このように、本発明の第3の実施の形態では、前述した第1の実施の形態の効果に加え管理端末300の機能をサーバ装置100に備えたので、サーバ装置とは別に管理端末を設けることなく、仮想計算機に対応した仮想ストレージの管理をすることができる。

【0118】

図15は、本発明の第4の実施の形態の計算機システムの構成を示すブロック図である。

【0119】

第4の実施の形態では、物理ストレージ制御部1100の構成が前述した実施の形態と異なる。物理ストレージ制御部1100は、サーバ装置100との通信を制御する一つ又は2以上のチャネルアダプタ1101と、物理ディスク215を制御する一つ又は2以上のディスクアダプタ1102と、一つ又は2以上のディスクキャッシュ1103と一つ又は2以上の制御CPU212が内部ネットワーク1104で結合されている。

【0120】

内部ネットワーク1104を持つ物理ストレージ制御部1100では、内部ネットワーク1104の帯域もストレージ装置200の性能に影響する重要な要因となる。そこで、本実施の形態では、ストレージハイパーバイザ220が内部ネットワーク1104の帯域を仮想計算機(0)131、仮想計算機(1)141毎に配分し、その配分に従って制御CPU212が入出力処理を実行する。帯域の制御方法には様々な方式が考えられるが、本発明はその制御方法に依存しない。

【0121】

さらに、仮想ディスク225の構成方法も性能に影響する要因である。前述したように仮想ディスク225は、物理ディスク215の記憶領域をストレージハイパーバイザ220がディスクとして仮想計算機(0)131、仮想計算機(1)141に見せるものである。従って、一つの仮想ディスク225の入出力性能を向上させたいときは、複数の物理ディスク215の一部の領域を抽出し、これら領域を合わせて仮想ディスク225を実現させればよい。なぜなら、多数の物理ディスク215が同時に並列に動作することによって、仮想計算機(0)131、仮想計算機(1)141の入出力要求を処理するからである。このことを図16を用いて詳細に説明する。

【0122】

図16(a)では、仮想ディスク1200は、一つの物理ディスク1201から構成されている。一方、図16(b)では、仮想ディスク1202は、三つの物理ディスク1203、1204、1205から、それぞれ一部の領域を抽出して構成されている。物理ディスク1201の性能は、単位時間内に実行できる入出力処理回数で示すことができる。これをxとすると仮想ディスク1200の入出力性能は、xである。一方、仮想ディスク1202の性能は、仮想計算機(0)131、仮想計算機(1)141が仮想ディスク1202の全領域に均等にアクセスすると仮定すると、物理ディスク1203、1204、1205は同時に並行して動作できるため、3xである。このように、仮想ディスク1202を構成する物理ディスク215の数によって性能が大きく影響される。

【0123】

従って、仮想ディスク225を構成するために使用される物理ディスク215の数も仮想計算機(0)131、仮想計算機(1)141のアプリケーション等にあわせて管理端末300から指定できるとよい。たとえば、ディスク上に広範囲にランダムアクセスするアプリケーションを仮想計算機(0)131、仮想計算機(1)141が実行している場合には、前述したようにディスクキャッシュ214の効果が生じることは少ない。このような場合には、物理ディスク215のアクセス性能が仮想ディスク225の性能を定める支配要因になるため、仮想ディスク225を構成する物理ディスク215の数を多くすることによって、仮想ディスク225の性能を向上させることができる。

【0124】

また、制御CPU212も、ストレージ装置200の入出力性能を決定する要因の一つである。制御CPU212の配分率も、仮想計算機(0)131、仮想計算機(1)141に必要とされる入出力性能や、アプリケーションに合わせて使用者が管理端末300より指定できるようにするとよい。ストレージ装置200の実現方法によっては、チャンネルアダプタ1101、ディスクアダプタ1102にそれぞれに制御CPU212が搭載された形態も考えられる。この場合、それぞれの仮想計算機(0)131、仮想計算機(1)141に関するデータの入出力を担当するチャンネルアダプタ1101、ディスクアダプタ1102を管理端末から指定することになる。

【0125】

以上で述べた、ストレージ装置200の資源(内部ネットワーク1104、物理ディスク215、制御CPU212等)を管理端末300から設定可能とするため、ストレージ資源管理テーブル223を変更する必要がある。

【0126】

図17は、本発明の第4の実施の形態のストレージ資源管理テーブル223の説明図である。

【0127】

図17では、ストレージ資源管理テーブル223に内部帯域1300の設定欄が追加されている。これは、仮想計算機(0)131、仮想計算機(1)141に割り当てる内部ネットワーク1104の帯域を指定する欄である。本実施の形態では、全帯域に占める割合を指定している。制御CPU212は、仮想計算機(0)131、仮想計算機(1)141が使用する内部ネットワーク帯域を監視し、入出力処理を遅らせることによって内部ネットワーク帯域を設定値を超えないように制御する。

【0128】

制御CPUの割り当ては、ストレージ資源管理テーブル223の制御CPU欄604によって設定できる。本欄にて仮想計算機(0)131、仮想計算機(1)141の入出力を担当する制御CPU212を指定する。割り当てられた制御CPU212の数が多ければ、それだけ入出力性能も高いことが期待できる。一つの制御CPU212が複数の仮想計算機(0)131、仮想計算機(1)141の入出力を担当することも可能である。この場合、当該制御CPU212が各仮想計算機が使用するCPU時間を監視し、仮想計算機(0)131、仮想計算機(1)141とのCPU時間配分を制御する。

【0129】

物理ディスクの配分は、仮想ディスク管理テーブル221によって管理されるため、前述した実施の形態と同様である。

【0130】

図18は、本発明の第4の実施の形態の計算機システムの構成の設定画面の説明図である。

【0131】

「CPU」欄は、担当する制御CPU212の数を指定する欄である。「ディスクキャッシュ」欄は、仮想計算機(0)131に割り当てられたディスクキャッシュの容量である。「内部帯域」欄は、仮想計算機(0)131に割り当てられたストレージ装置200の内部ネットワーク1104の帯域である。「ディスク」欄は、仮想計算機(0)131に割り当てた論理ユニット231の個数を表示している。「ディスク」欄に設けられた「詳細」ボタンを押すことによって、各論理ユニットの物理ディスクの構成を表示し、設定する画面(図12)を表示することができる。

【図面の簡単な説明】

【0132】

【図1】本発明の第1の実施の形態の計算機システムの構成を示すブロック図である。

【図2】本発明の実施の形態の仮想ディスク管理テーブルの説明図である。

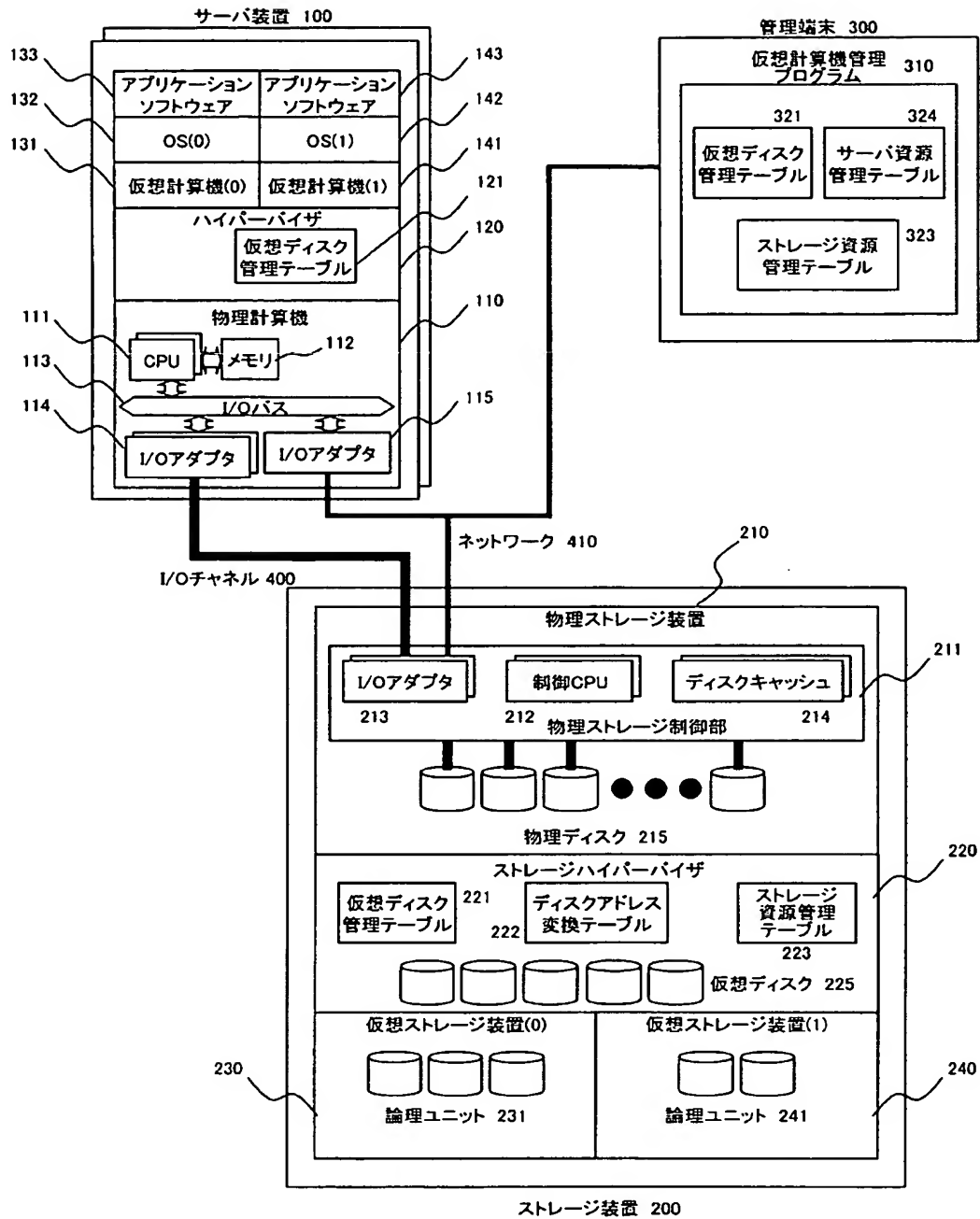
- 【図 3】 本発明の実施の形態のディスクアドレス変換テーブルの説明図である。
- 【図 4】 本発明の実施の形態のストレージ資源管理テーブルの説明図である。
- 【図 5】 本発明の実施の形態の資源管理テーブルの説明図である。
- 【図 6】 本発明の実施の形態の資源割り当て設定処理のフローチャートである。
- 【図 7】 本発明の実施の形態のデータ入出力処理のフローチャートである。
- 【図 8】 本発明の実施の形態の I/O チャンネルの通信プロトコル層の構造の説明図である。
- 【図 9】 本発明の実施の形態のサーバ装置とストレージ装置との間で伝達される情報の説明図である。
- 【図 10】 本発明の実施の形態のハイパーバイザ通信ヘッダ 1203 の内容の説明図である。
- 【図 11】 本発明の実施の形態の計算機システムの構成の設定画面の説明図である。
- 【図 12】 本発明の実施の形態の計算機システムの構成の設定画面の説明図である。
- 【図 13】 本発明の第 2 の実施の形態の計算機システムの構成を示すブロック図である。
- 【図 14】 本発明の第 3 の実施の形態の計算機システムの構成を示すブロック図である。
- 【図 15】 本発明の第 4 の実施の形態の計算機システムの構成を示すブロック図である。
- 【図 16】 本発明の第 4 の実施の形態の物理ディスクの動作の説明図である。
- 【図 17】 本発明の第 4 の実施の形態のストレージ資源管理テーブルの説明図である。
- 【図 18】 本発明の第 4 の実施の形態の計算機システムの構成の設定画面の説明図である。

【符号の説明】

【0133】

- 100 サーバ装置
- 110 物理計算機
- 120 ハイパーバイザ
- 121、221、321 仮想ディスク管理テーブル
- 123、223、323 ストレージ資源管理テーブル
- 124、224、324 サーバ資源管理テーブル
- 131、141 仮想計算機
- 200 ストレージ装置
- 210 物理ストレージ装置
- 220 ストレージハイパーバイザ
- 222 ディスクアドレス管理テーブル
- 230、240 仮想ストレージ装置
- 300、350 管理端末
- 400 I/O チャンネル
- 410 ネットワーク

【書類名】 図面
【図 1】



【図 2】

仮想計算機番号	論理ユニット番号	仮想ディスク番号	仮想ストレージ装置番号
0	0	121	0
	1	122	
	2	123	
1	3	16	1
	4	17	
	5	18	
⋮	⋮	⋮	

仮想ディスク管理テーブル

221,121,321

【図 3】

仮想ディスク番号	仮想ブロック アドレス	物理ディスク番号	物理ブロック アドレス
121	0x00000000	8	0x00000000
	0x80000000	9	0x00000000
122	0x00000000	10	0x10000000
⋮	⋮	⋮	⋮

ディスクアドレス変換テーブル

222

【図 4】

仮想計算機番号	仮想ディスク番号	ディスクキャッシュ容量	制御CPU	I/Oアダプタ
0	121	512MB	48	0
	122		49	1
	123		50	2
1	16	256MB	112	3
	17		113	4
	18			
● ● ●	● ● ●	● ● ●	● ● ●	● ● ●

ストレージ資源管理テーブル

323,223,123

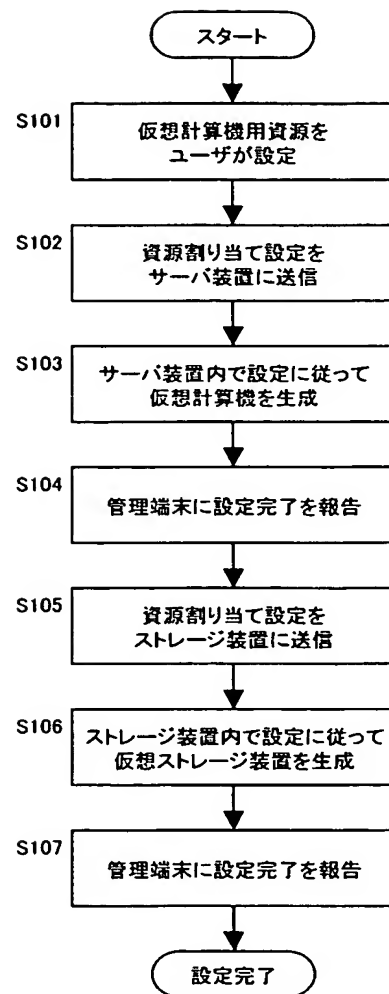
【図 5】

仮想計算機番号	CPU配分率	メモリ容量	I/Oアダプタ
0	20%	512MB	0 1 2
1	30%	128MB	3 4
⋮	⋮	⋮	⋮

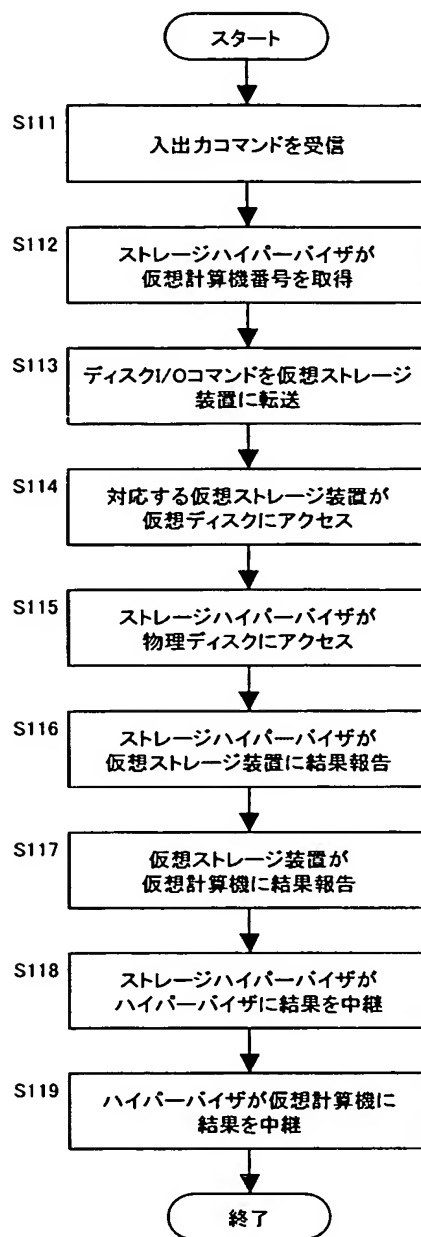
サーバ資源管理テーブル

324,124,224

【図 6】

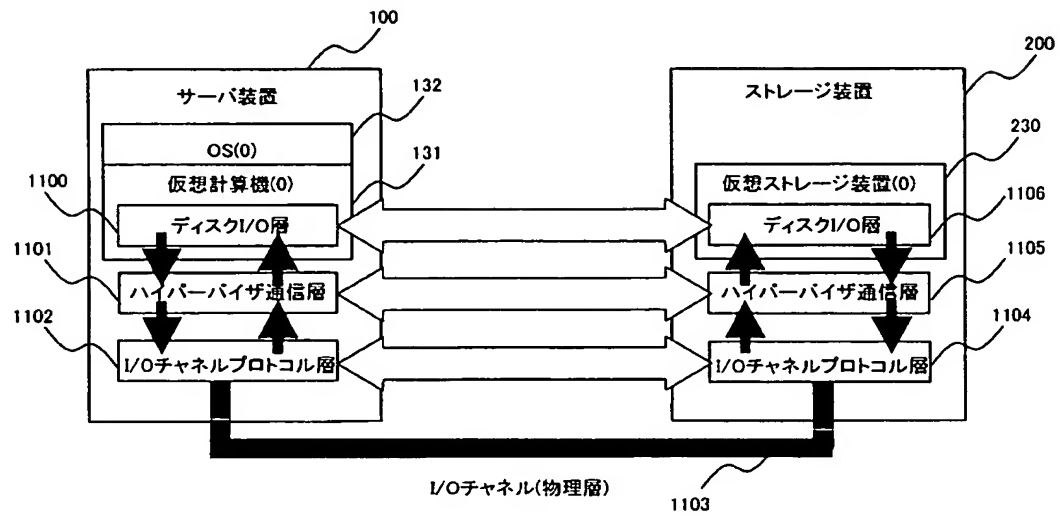


【図 7】

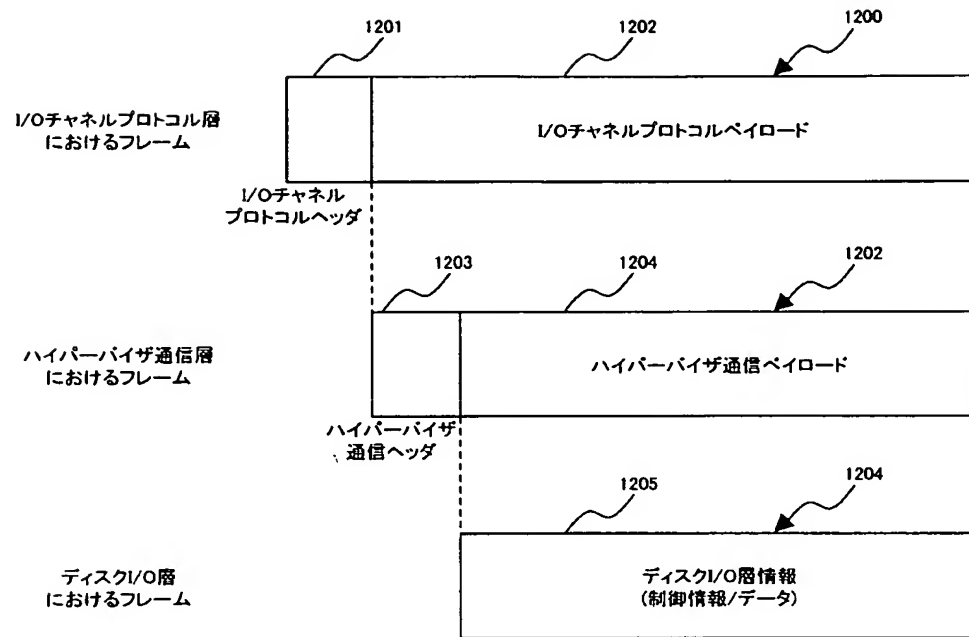


【図 8】

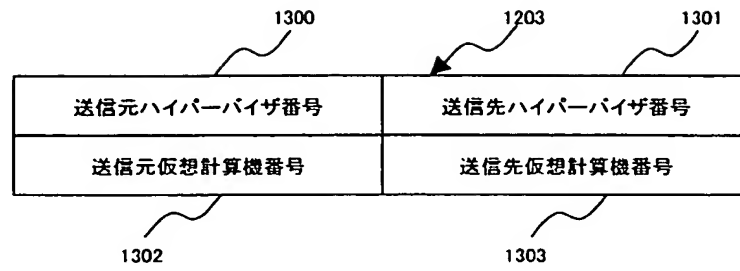
BEST AVAILABLE COPY



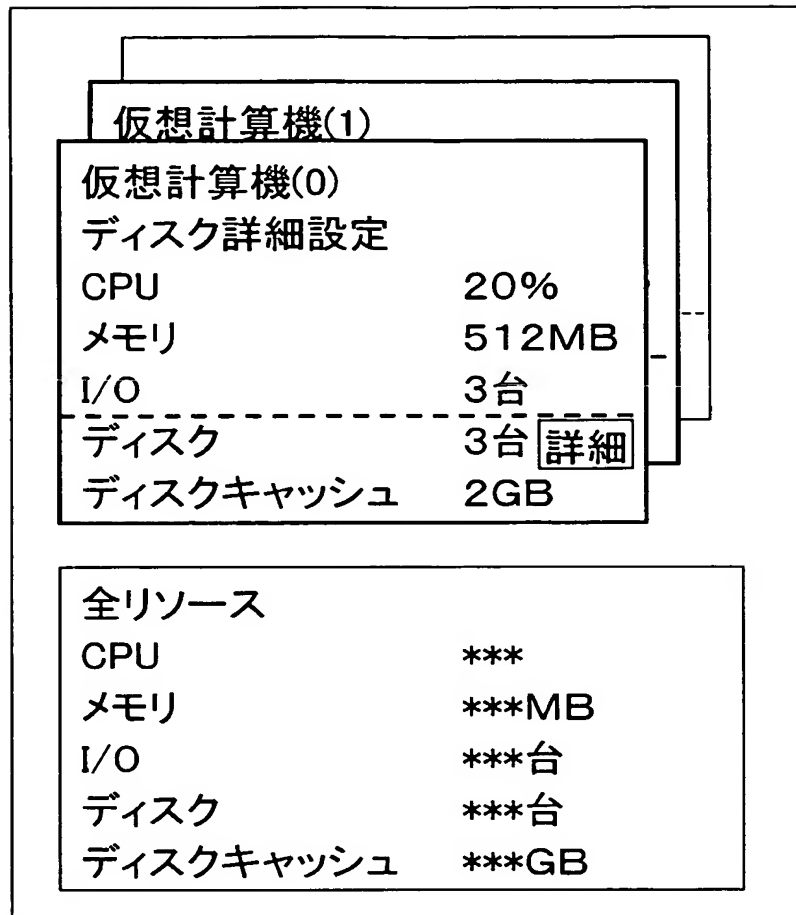
【図 9】



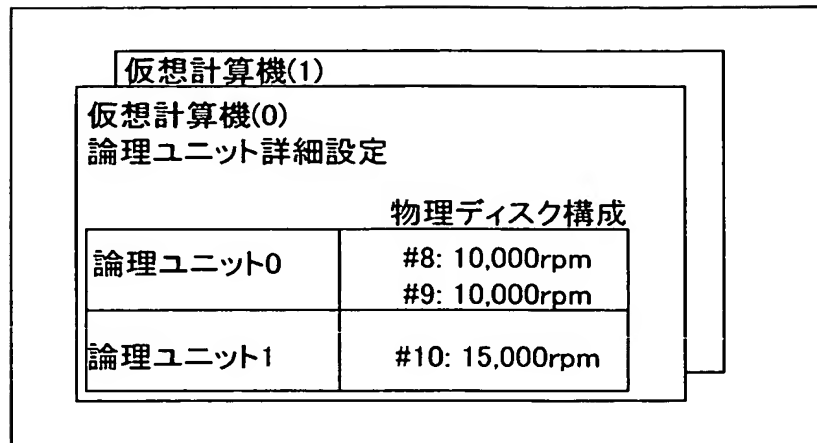
【図 10】



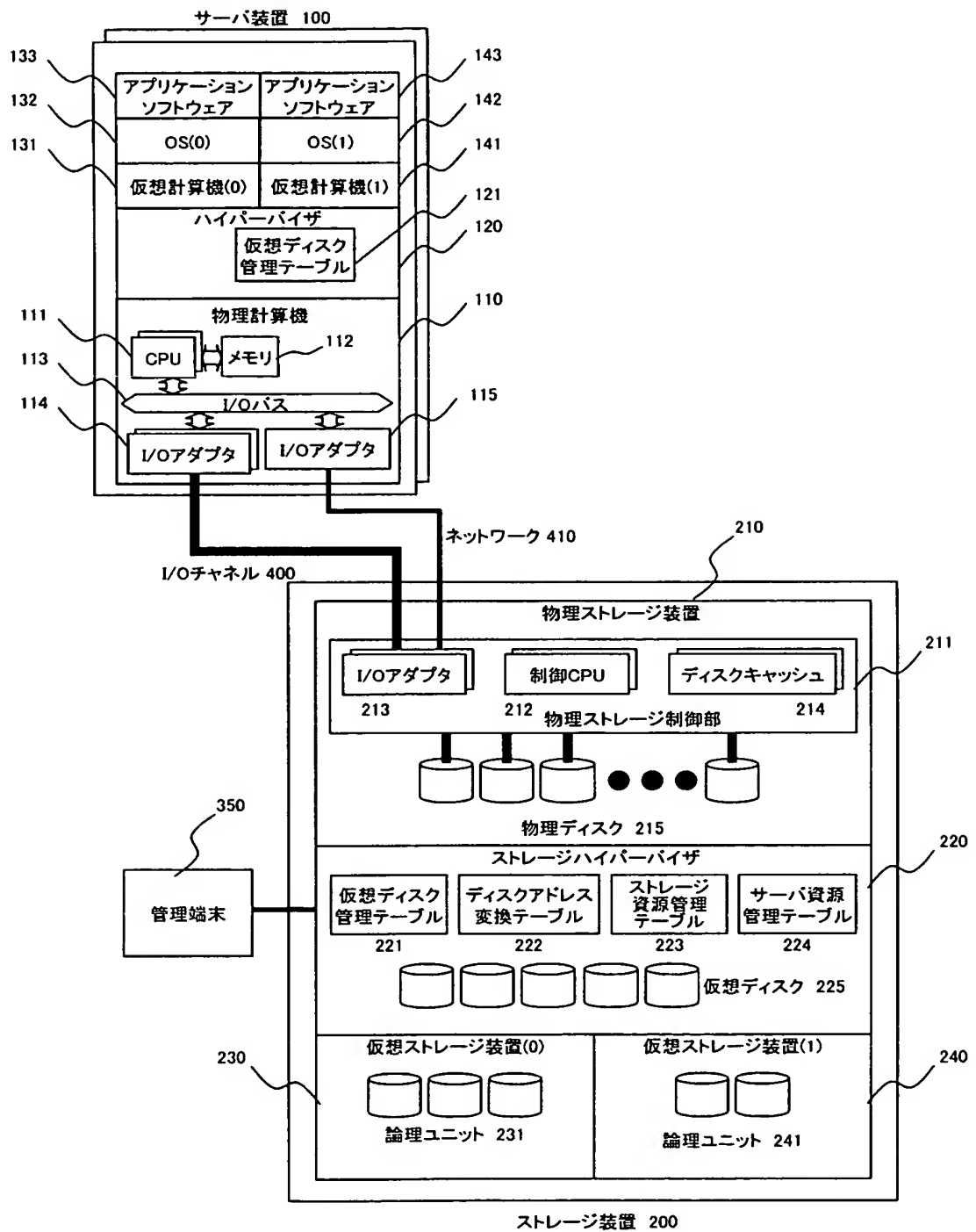
【図 11】



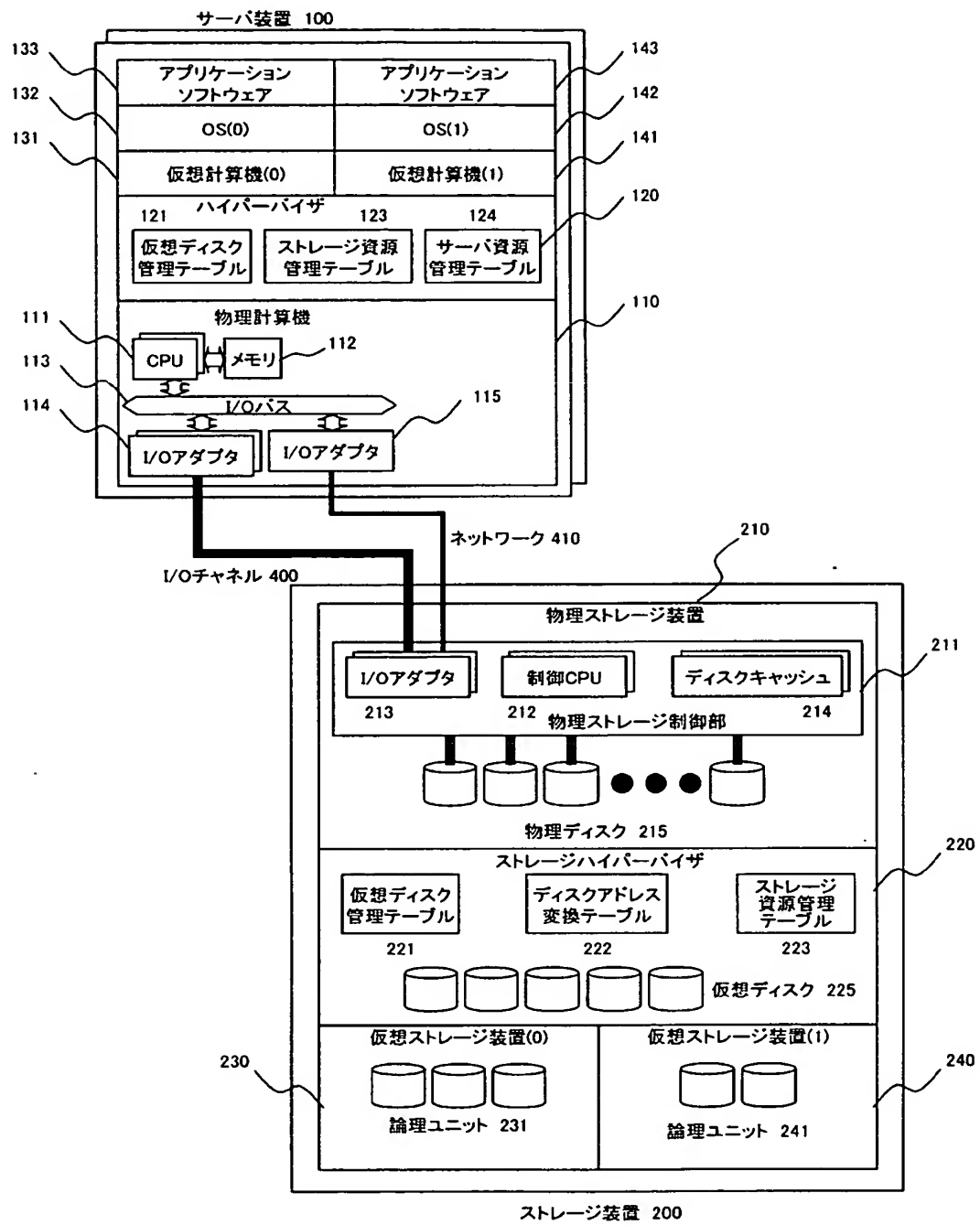
【図 12】



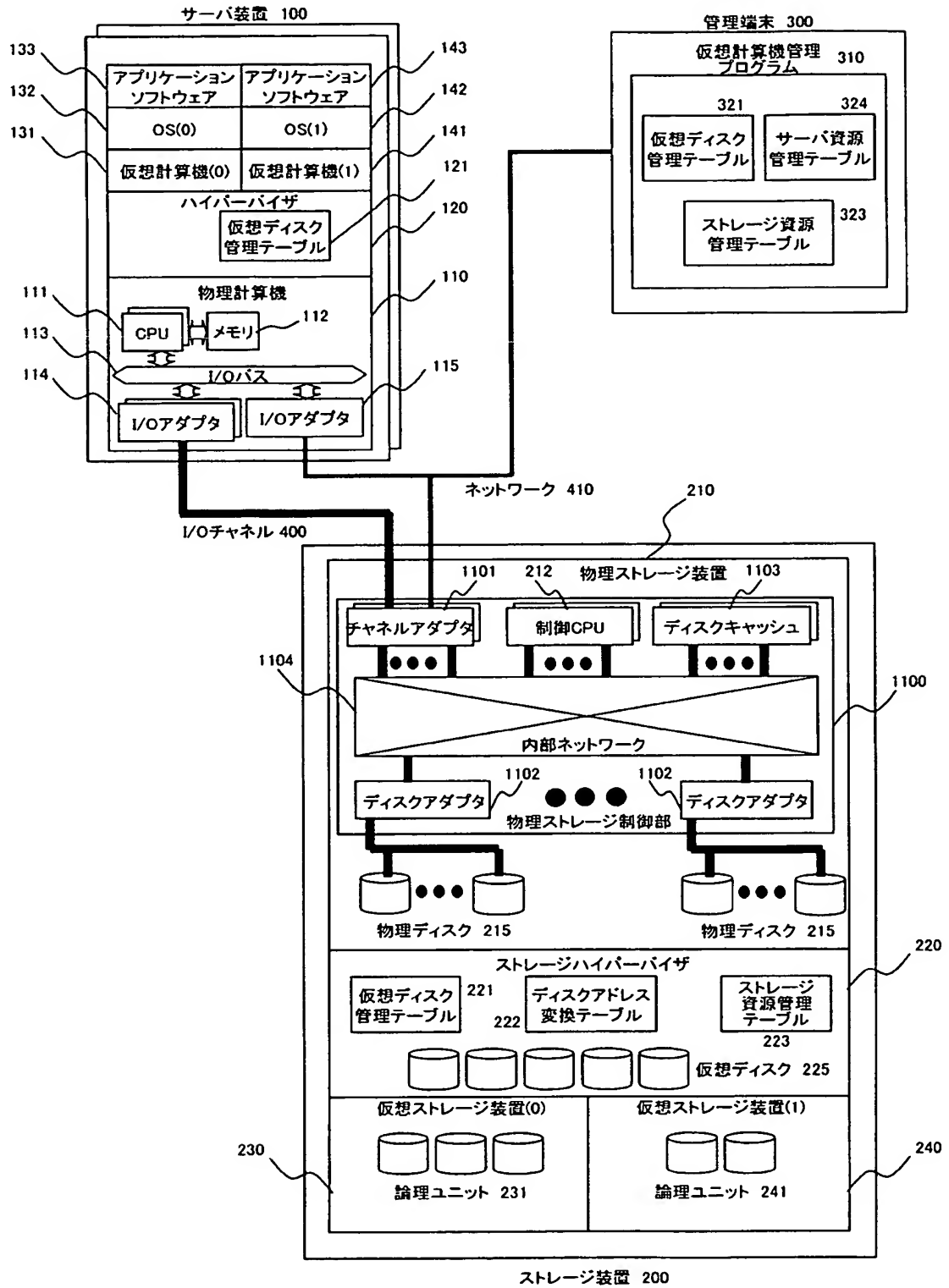
【図 13】



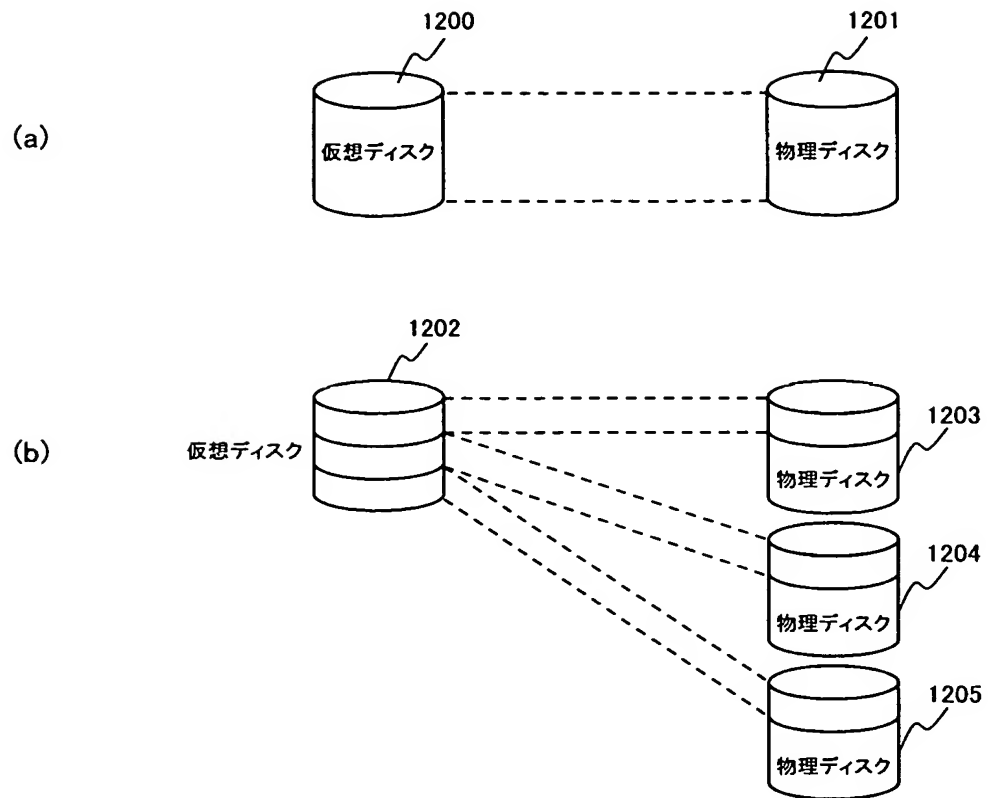
【図 14】



【図 15】



【図 16】



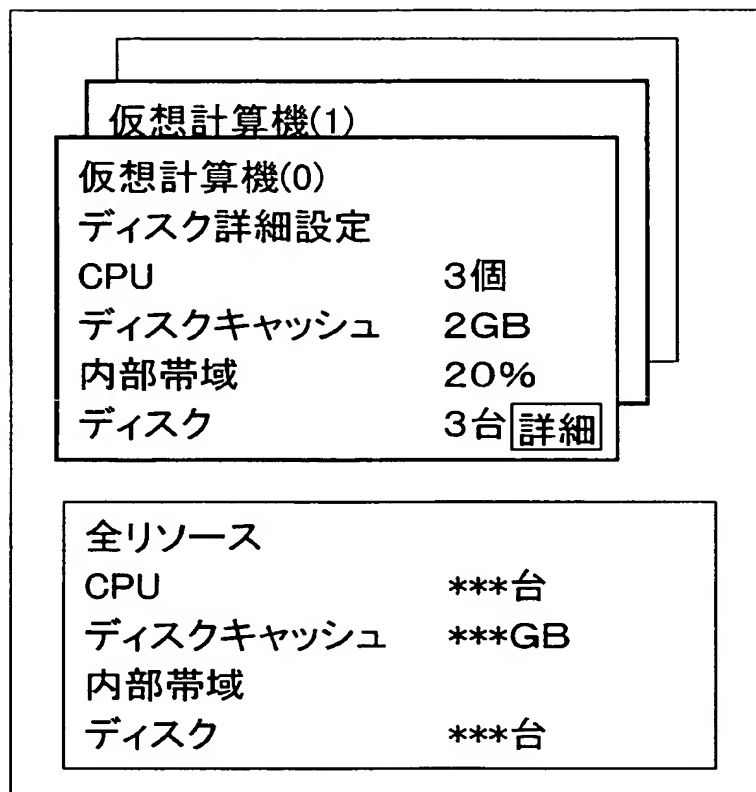
【図 17】

仮想計算機番号	仮想ディスク番号	ディスクキャッシュ容量	制御CPU	チャネルアダプタ	内部帯域
0	121	512MB	48	0	20%
	122		49	1	
	123		50	2	
1	16	256MB	112	3	10%
	17		113	4	
	18				
⋮	⋮	⋮	⋮	⋮	

ストレージ資源管理テーブル

323.223

【図 18】



【書類名】 要約書**【要約】**

【課題】複数のホストによって共有されるストレージ装置の利用率を高め、ホストとストレージを含めたシステム全体での性能の最適化を図る。

【解決手段】コンピュータ装置は、該コンピュータ装置に備わる計算資源を論理的に分割し、独立した仮想計算機として動作させる第1制御部を有し；ストレージ装置は、該ストレージ装置に備わるストレージ資源を論理的に分割し、独立した仮想ストレージ装置として動作させる第2制御部を有し；コンピュータ装置に備わる計算資源を管理する第1管理テーブルと、ストレージ装置に備わるストレージ資源を管理する第2管理テーブルと、仮想計算機と仮想ストレージ装置との関係を管理する第3管理テーブルと、を有する管理部を備え；第1制御部は、第1管理テーブルの設定に基づいて計算資源を論理的に分割し；第2制御部は、第2管理テーブルの設定に基づいてストレージ資源を論理的に分割する。

【効果】

【選択図】 図1

特願 2 0 0 4 - 0 2 6 5 7 5

出 願 人 履 歴 情 報

識別番号 [0 0 0 0 0 5 1 0 8]

1. 変更年月日	1 9 9 0 年 8 月 3 1 日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台 4 丁目 6 番地
氏 名	株式会社日立製作所